Wavelet Domain Geophysical Inversion

Jonathan Kane, Felix Herrmann, and M. Nafi Toksöz
Earth Resources Laboratory
Dept. of Earth, Atmospheric, and Planetary Sciences
Massachusetts Institute of Technology
Cambridge, MA 02139

Abstract

We present a non-linear method for solving linear inverse problems by thresholding coefficients in the wavelet domain[1]. Our method is based on the wavelet-vaguelette decomposition of Donoho (1992).

Numerical results for a synthetic travel-time inversion problem show that the wavelet based method outperforms traditional least-squares methods of solution.

1 Introduction

Recently wavelets have established themselves as highly useful mathematical functions for a variety of applications such as data compression and denoising of audio and video signals. Denoising an observed signal can be looked upon as an inverse problem, where the forward modeling operator is simply the identity operator on the model space. Given the success of wavelets in this application, it is reasonable to ask whether they are applicable to other inverse problems. In the 1990's a large amount of work in the applied mathematics and statistics community has gone into answering this question. Perhaps foremost among the workers have been Donoho (1992), who formulated the wavelet-vaguelette paradigm for solving certain linear inverse problems. Their theory has been applied by various workers: Abramovich and Silverman (1998), formulated a related vaguelette-wavelet method; Kolaczyk (1996) applied the method to 2-D tomographic problems; and Neelami et al. (1999) to deconvolution problems.

In this work we apply the wavelet-vaguelette method to a synthetic geophysical inverse problem, that of slowness estimation from travel-time data. Li et al. (1996) previously used wavelets to solve this inverse problem although not by the wavelet-vaguelette method. We incorporate a new class of wavelets in the inversion: **fractional spline wavelets**. These recent additions to the wavelet family were created by Unser and Blu (1999) and are generalizations of the Battle-Lemarié wavelets. They have an adjustable "smoothness" parameter, α that allows the wavelets to be better fit to a problem at hand.

2 Fractional Splines

Following Unser and Blu (1999) we define an α degree causal **B-spline**, β_+^α, as multiple convolutions of the boxcar function with itself. Multiple auto-convolutions of this functions are best expressed in the Fourier domain as multiplications:

$$\hat{\beta}_+^\alpha(\omega) = \left(\frac{1-e^{-i\omega}}{i\omega}\right)^{\alpha+1}. \tag{1}$$

With traditional B-splines α was restricted to be an integer. Unser and Blu (1999) have relaxed this restriction and the resulting functions for $\alpha \in \{\mathbb{R} : \alpha > -1/2\}$ are known as **fractional B-splines**. Examples of these functions are shown in Figure 1 for varying degree α.

[1] For an alternative, data adaptive, approach to using wavelets and thresholding in solving inverse problems see Herrmann et al. in these proceedings.

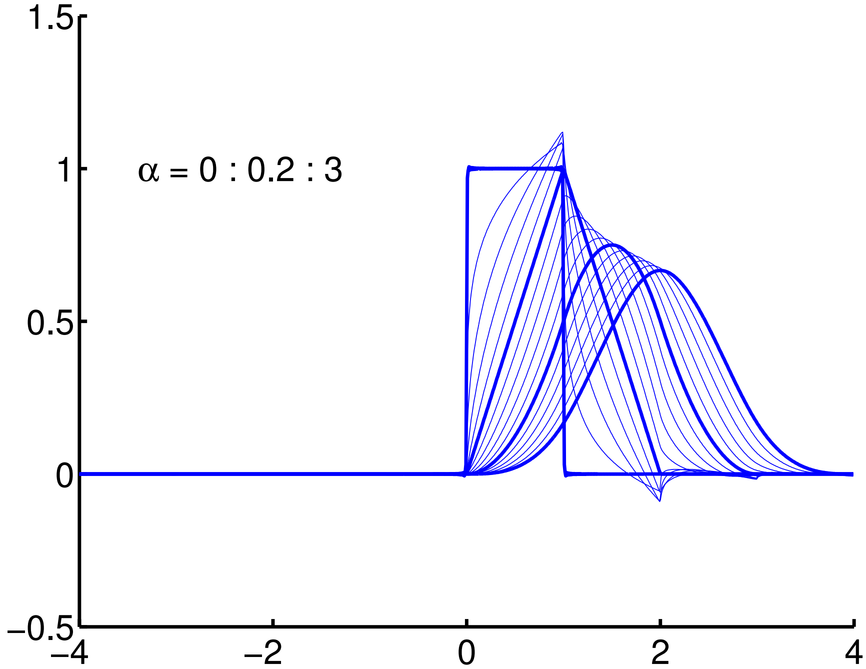

Figure 1: Causal B-splines

B-splines can be converted into orthogonal splines via the following equation:

$$\hat{\gamma}^{\alpha}(\omega) = \frac{\hat{\beta}^{\alpha}(\omega)}{\left(\sum_k |\hat{\beta}^{\alpha}(\omega + 2\pi k)|^2\right)^{1/2}}, \tag{2}$$

which is simply dividing the B-spline in the Fourier domain by the square root 2π-periodic power spectrum. These functions have the additional useful property of being orthogonal to integer shifts of themselves. Using orthogonal splines as scaling functions leads to an orthogonal wavelet transform.

3 Wavelets

Wavelets are localized oscillatory functions, i.e. “wiggles” that have most of their energy at a specific point in space. They must integrate to zero to be a proper wavelet, and, depending on the wavelet, have higher order moments also equal to zero. Wavelets are contructed from a **scaling function**, a basis function that does *not* integrate to zero. For an arbitrary scaling function, $\phi(x)$, satisfying the necessary properties put forth by Mallat (1998), we can construct a **refinement filter** (here presented in the Fourier domain):

$$\hat{h}(\omega) = \sqrt{2}\,\frac{\hat{\phi}(2\omega)}{\hat{\phi}(\omega)}, \tag{3}$$

which is a low-pass filter. This in turn leads to a high-pass filter via the equation presented in Mallat (1998):

$$\hat{g}(\omega) = -e^{-i\omega}\hat{h}^{*}(\omega + \pi), \tag{4}$$

which then leads to the actual wavelet

$$\hat{\psi}(\omega) = \frac{1}{\sqrt{2}}\,\hat{g}(\frac{\omega}{2})\hat{\phi}(\frac{\omega}{2}). \tag{5}$$

In this work we use orthogonal fractional splines as scaling functions and the wavelets generated by equation 5 are **orthogonal fractional spline wavelets**. We show orthogonal fractional spline wavelets for varying degrees of α in Figure 2.

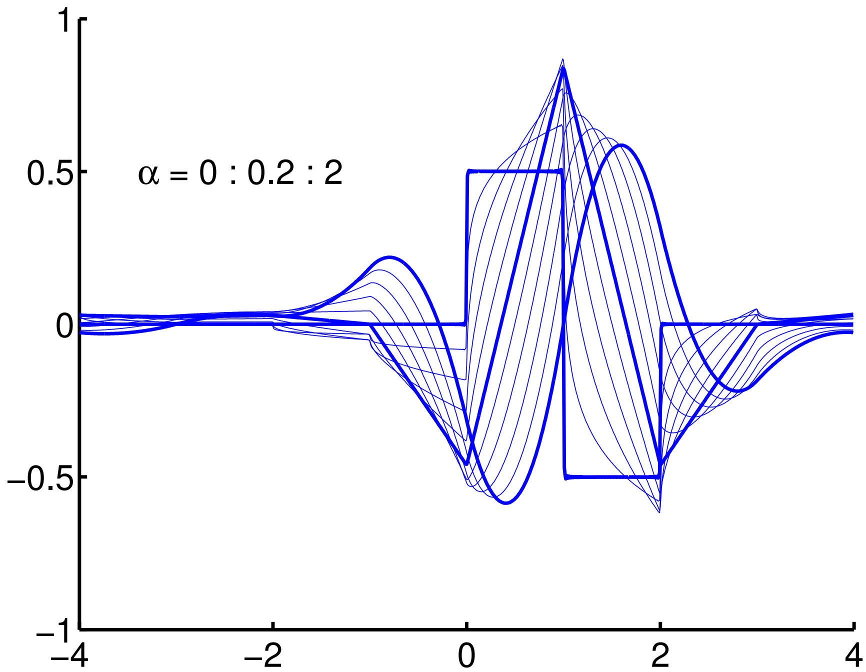

Figure 2: Causal orthogonal splines wavelets for α from 0 to 2

The discrete wavelet transform (DWT) analyzes a continuous function, fx, by inner products with translations of a dilated scaling function, $\phi(\mathbf{x})$, and translated and dilated wavelets, $\psi(\mathbf{x})$:

$$\begin{aligned} \mathbf{u} = u_{J,\mathbf{k}} &= \int fx \left[\frac{1}{\sqrt{2^J}} \phi(\frac{\mathbf{x}}{2^J} - \mathbf{k}) \right]^* dx, \\ \mathbf{v} = v_{j,\mathbf{k}} &= \int fx \left[\frac{1}{\sqrt{2^j}} \psi(\frac{\mathbf{x}}{2^j} - \mathbf{k}) \right]^* dx, \end{aligned} \tag{6}$$

where $\mathbf{u}$ and $\mathbf{v}$ are vectors of discrete wavelet coefficients.

If ϕ, ψ, and f are discretized, we can put the transformation in equation 6 into a matrix formulation:

$$\tilde{\mathbf{f}} = \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} = \mathbf{W}\mathbf{f}, \tag{7}$$

where $\tilde{\mathbf{f}}$ are the wavelet coefficients. In the special case that orthogonal wavelets are used, $\mathbf{W}$ is an orthogonal matrix.

4 Linear Inverse Theory

Linear inverse problems are often encountered in science and mathematics and take the form

$$\mathbf{z} = \mathbf{K}\mathbf{f} + \mathbf{n}. \tag{8}$$

where $\mathbf{f}$ is a function we desire to estimate, $\mathbf{n}$ is the uncertainty (noise) in the system (which is unrelated to $\mathbf{f}$), $\mathbf{K}$ is a linear operator mapping models into data, and $\mathbf{z}$ is the noise contaminated data.

A solution to this problem is obtained by minimizing

$$\|\mathbf{z} - \mathbf{K}\mathbf{f}\|_{\mathcal{L}_2}. \tag{9}$$

If $\mathbf{K}$ is rectangular and $\mathbf{K}^T\mathbf{K}$ is invertible, this leads to

$$\mathbf{f}_{est} = (\mathbf{K}^T\mathbf{K})^{-1}\mathbf{K}^T\mathbf{z}. \tag{10}$$

If $\mathbf{K}$ is square and invertible we have

$$\mathbf{f}_{est} = \mathbf{K}^{-1}\mathbf{z}. \tag{11}$$

If we substitute equation 8 into equation 11 we obtain

$$\mathbf{f}_{est} = \mathbf{f} + \mathbf{K}^{-1}\mathbf{n}. \tag{12}$$

From equation 12 we can see that our solution will be contaminated with colored noise. If this term is of too large a magnitude, or if $\mathbf{K}$ was not invertible in the first place, we have an **ill-posed** inverse problem. The traditional way to solve such a problem is with **regularization** where we redefine the minimization problem 9 by adding one of the following constraints:

$$\min_{\mathbf{f}}\|\mathbf{f}\|_{\mathcal{L}_2}, \qquad \text{or} \qquad \min_{\mathbf{f}}\|\mathbf{L}\mathbf{f}\|_{\mathcal{L}_2}, \tag{13}$$

where $\mathbf{L}$ is usually a differential operator. The first of these constraints leads to the **damped least squares** solution:

$$\mathbf{f}_{est} = \underbrace{(\mathbf{K}^T\mathbf{K} + \zeta\,\mathbf{I})^{-1}\mathbf{K}^T}_{\mathbf{K}_{dls}^{-1}}\mathbf{z}, \tag{14}$$

The second constraint in equation 13 leads to the **generalized least squares** solution:

$$\mathbf{f}_{est} = \underbrace{(\mathbf{K}^T\mathbf{K} + \zeta\,\mathbf{L}^T\mathbf{L})^{-1}\mathbf{K}^T}_{\mathbf{K}_{gls}^{-1}}\mathbf{z}. \tag{15}$$

ζ is an adjustable constant that changes the amount of noise damping in the inversion. ζ is hard to estimate in advance and a poor choice of it can lead to poor results in the inversion. If it is too large, the solution will be too smooth. If it is too small the solution will be too rough.

5 The Wavelet-Vaguelette Decomposition

Donoho (1992) presented an alternative methodology for solving a certain set of linear inverse problems in which $\mathbf{K}$ is a **homogeneous operator**. Homogeneous operators include differentiation, radon transforms, and, important to this work, integration operators. In a matrix formulation, the method proceeds as follows: an orthogonal wavelet transform matrix $\mathbf{W}$ is constructed, where each row is a discrete wavelet. $\mathbf{K}$ then operates on each individual wavelet to produce what is called a **vaguelette**:

$$\mathbf{K}\mathbf{W}^T = \mathbf{V}^T\mathbf{\Gamma}. \tag{16}$$

Each column of the matrix $\mathbf{V}^T$ is a discrete vaguelette and has been normalized to unit energy. Each normalization factor has been put on the diagonal of the diagonal matrix $\mathbf{\Gamma}$. Moving $\mathbf{W}$ to the other side of equation 16, we get the **wavelet-vaguelette decomposition** (WVD):

$$\mathbf{K} = \mathbf{V}^T\mathbf{\Gamma}\mathbf{W}. \tag{17}$$

Equation 17 looks alot like the **singular value decomposition** (Strang, 1986), with the exception that $\mathbf{V}^T$ is not an orthogonal matrix (although its columns have unit energy and for homogeneous operators it is *near* orthogonal). Donoho (1992) therefore called the entries of $\mathbf{\Gamma}$ *quasi*-singular values. It is important to note here that $\mathbf{W}$ and $\mathbf{\Gamma}$ are always invertible, but $\mathbf{V}^T$ is invertible only if $\mathbf{K}$ is. If $\mathbf{V}^T$ is indeed invertible we call its inverse $\mathbf{U}$, which has as its rows the biorthogonal duals of the columns of $\mathbf{V}^T$.

We are now ready to apply the WVD to the solution of a linear inverse problem. Assuming that $\mathbf{K}$ is invertible and square we can represent its inverse via the WVD as

$$\mathbf{K}^{-1} = \mathbf{W}^T \mathbf{\Gamma}^{-1} \mathbf{U}. \tag{18}$$

Plugging equation 18 into equation 11 we get

$$\mathbf{f}_{est} = \mathbf{W}^T \mathbf{\Gamma}^{-1} \mathbf{U} \mathbf{z}. \tag{19}$$

Noting that $\mathbf{U} = \mathbf{\Gamma W K}^{-1}$, we substitute into equation 19 and obtain

$$\mathbf{f}_{est} = \mathbf{W}^T \mathbf{\Gamma}^{-1} \mathbf{\Gamma W K}^{-1} \mathbf{z}. \tag{20}$$

Looking at equation 20 we see that we have not yet really changed anything from equation 11; the first four matrices cancel themselves. This is where the crux of the WVD inversion method appears. We insert a non-linear thresholding operator, Θ_{T}, into the equation:

$$\mathbf{f}_{est} = \mathbf{W}^T \mathbf{\Gamma}^{-1} \Theta_{\mathrm{T}}[\mathbf{\Gamma W K}^{-1} \mathbf{z}]. \tag{21}$$

We define Θ_{T} by

$$\Theta_{\mathrm{T}}[\cdot] = \begin{cases} \mathbf{f}_j & : \quad |\mathbf{f}_j| > \mathrm{T} \\ 0 & : \quad |\mathbf{f}_j| \leq \mathrm{T} \end{cases}, \tag{22}$$

where a "universal" threshold criterion has been defined by Donoho (1992) as

$$\mathrm{T} = \sigma_{\mathbf{n}} \sqrt{2ln(\mathrm{N})}. \tag{23}$$

$\sigma_{\mathbf{n}}$ is the assumed standard deviation of the noise and N is the number of coefficients in $\mathbf{f}$.

The intuitive idea here is this: solving equation 11 leads to a noise contaminated solution. Transforming this solution to the wavelet domain tends to isolate "good" signal into a few large valued, isolated coefficients, while the noise tends to be spread around equally with smaller energy. Thus thresholding the small wavelet coefficients will tend to remove the noise and leave the more interesting coherent features untouched.

If the noise level, $\sigma_{\mathbf{n}}$, is small and $\mathbf{K}$ is invertible, the WVD decomposition with thresholding provides an excellent way to "regularize" the solution without minimizing any additional constraints such as in 13. This is idealistic though. If the noise level is large or $\mathbf{K}$ is non-invertible some classical regularization must be imposed to obtain a decent solution. This is done by substituting $\mathbf{K}_{gls}^{-1}$ or $\mathbf{K}_{dls}^{-1}$ instead of $\mathbf{K}^{-1}$ in equation 21. If ζ is kept very small the WVD thresholding method can outperform the optimal solution obtained *without* wavelet domain thresholding. We show this in the examples below. Doing such a *generalized* WVD thresholding method goes beyond the setting proscribed by Donoho (1992) and, to our knowledge, there are no theoretical results as to its optimality. However, numerical results show that it beats conventional least squares methods across a broad range of problems.

6 Travel-Time Inversion

In this section we apply the method to a slowness estimation problem to show its superiority to classical methods. In Figure 3 we show a slowness log measured in depth. This is the $\mathbf{f}$ that we will try to estimate. Our operator $\mathbf{K}$ is the integration operator. This will simulate having a series of receivers in a well measuring the time it takes for a pulse to travel down the well. The data is a monotonically increasing function measured in time. It is corrupted by white noise and shown in Figure 4. We choose an orthogonal fractional spline wavelet with $\alpha = 0.3492$ to perform the wavelet transform. We choose this particular α based on the method of Li et al. (1996) using the well log in Figure 3.

We compare results of inversion done in three ways: 1) by applying $\mathbf{K}^{-1}$ as in equation 11 (Figure 5), 2) by applying $\mathbf{K}_{dls}^{-1}$ according to equation 14 (Figure 6), and 3) applying $\mathbf{K}_{dls}^{-1}$ followed by WVD thresholding (Figure 7). The mean square error (MSE) is shown in each figure window.

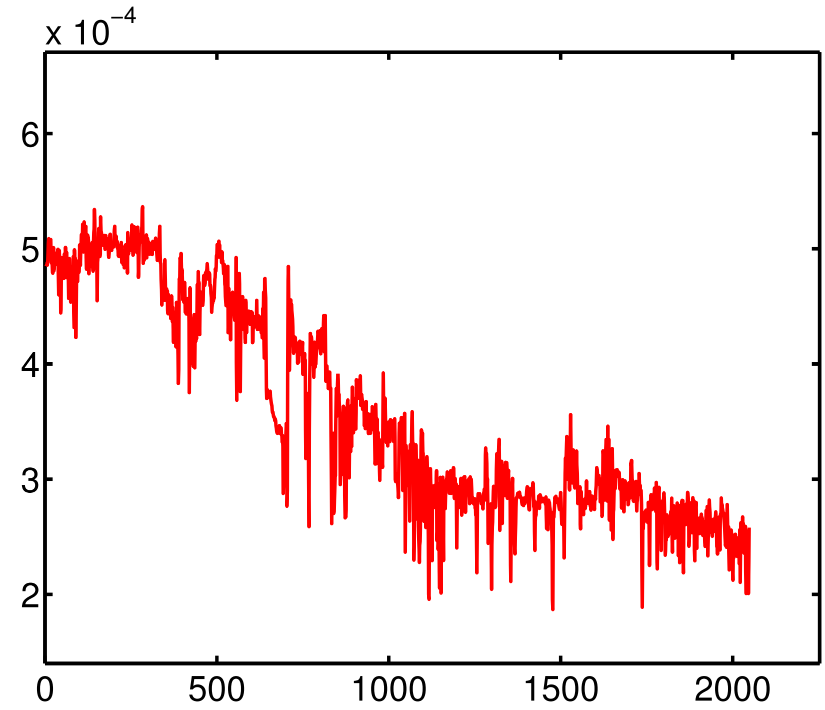

Figure 3: Slowness log

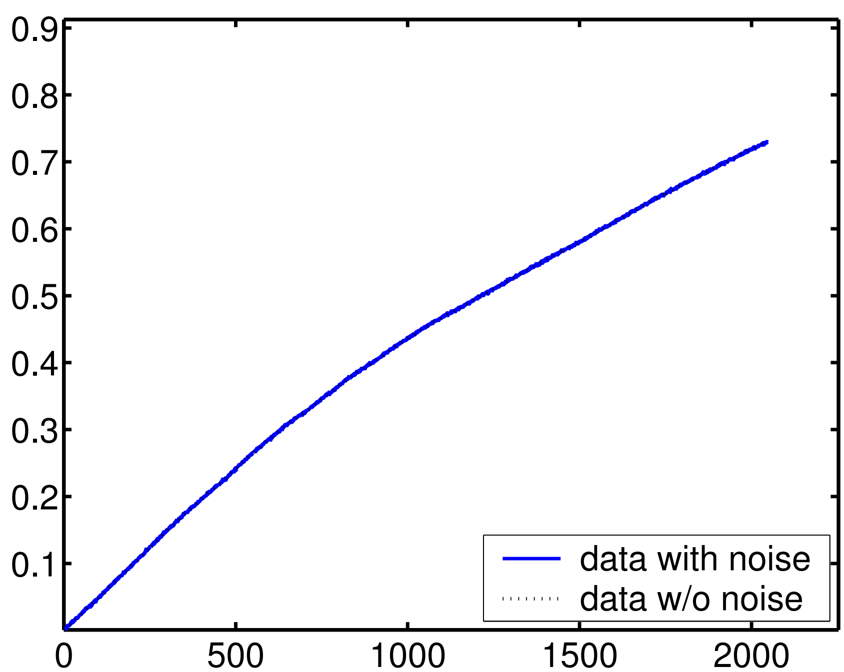

Figure 4: Integrated slowness (travel-time) data

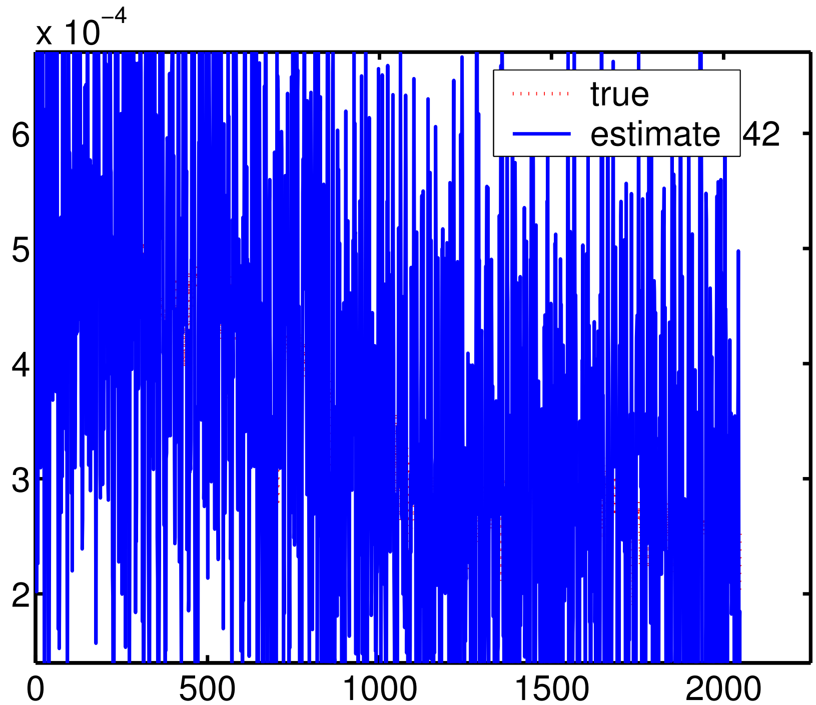

Figure 5: Inverse according to equation 11

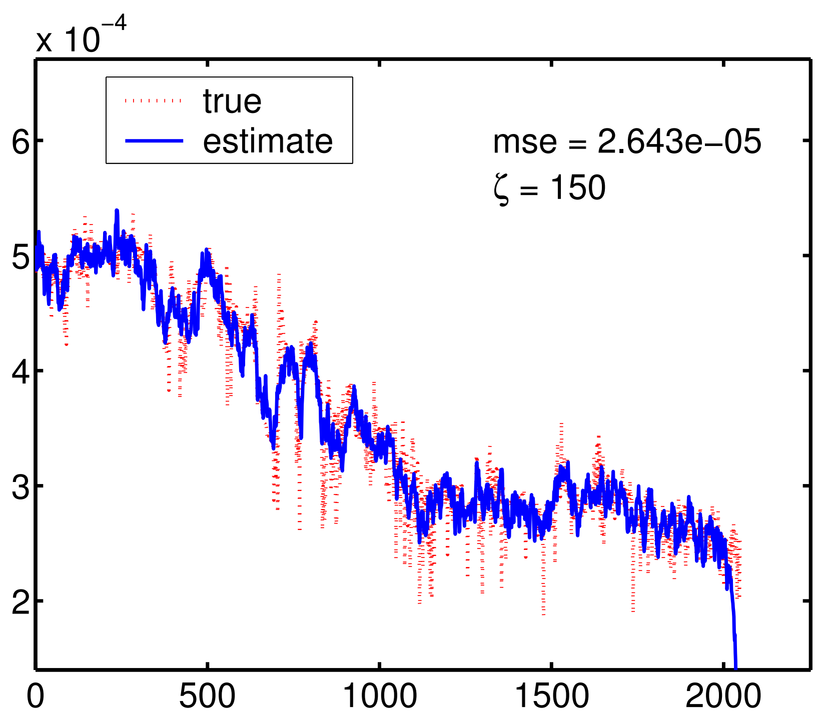

Figure 6: Damped least squares inverse

The results of inversion according shown in Figure 5 are highly contaminated with noise and useless. In Figure 6 we see a great improvement in the results, but much of the apparent detail at fine scales is actually noise and not accurate. The results in Figure 7 of damped least squares with subsequent WVD thresholding exhibit the smallest MSE and bring out edges in the data clearly. We have observed that no matter which

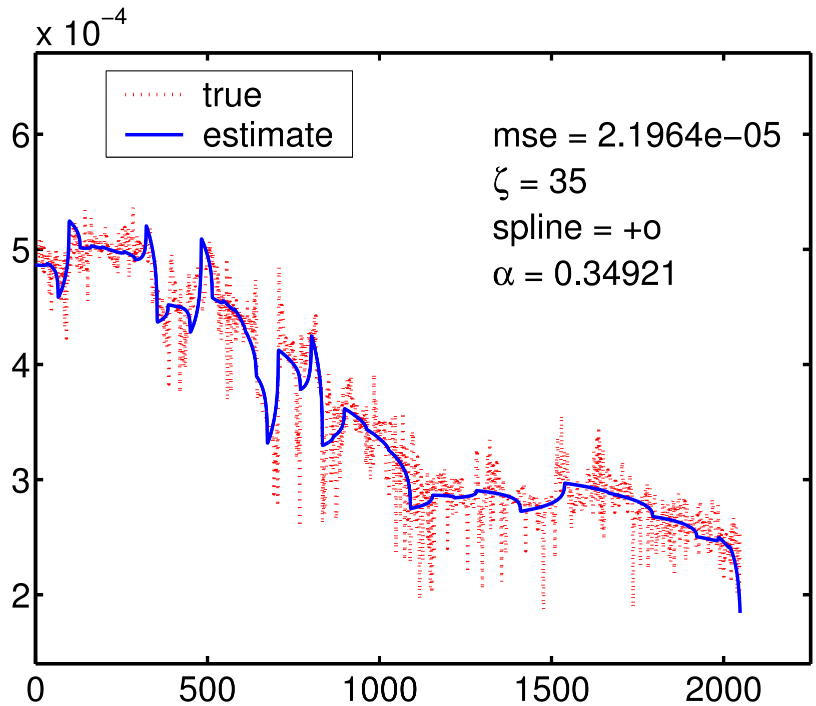

Figure 7: Damped least squares inverse with WVD thresholding

inverse method was used the MSE was consistently lower when WVD thresholding was applied than the when it was not. For each method the optimal ζ was chosen by trial and error.

7 Conclusions

We have demonstrated the WVD method for a one-dimensional linear inverse problem. The method was compared to the classical damped least-squares method. We see an improvement in MSE of WVD with thresholding over a classical damped least squares method with an optimally chosen damping parameter ζ. This result is consistently observed across a wide variety of models and operators.

8 Acknowledgments

We would like to thank the ERL Founding Members for financial sponsorhip of this work. We especially thank Chevron Texaco for their fellowship supporting Jonathan Kane.

References

Abramovich, F. and Silverman, B. W. (1998). Wavelet decomposition approaches to statistical inverse problems. *Biometrika*.

Donoho, D. L. (1992). Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. Technical report, Dept. of Statistics, Stanford University.

Kolaczyk, E. D. (1996). A wavelet shrinkage approach to tomographic image reconstruction. *Journal of the American Statistical Association*.

Li, X., Sacchi, M. D., and Ulrych, T. J. (1996). Wavelet transform inversion with prior scale information. *Geophysics*, 61(05):1379–1385.

Mallat, S. G. (1998). *A Wavelet Tour of Signal Processing*. Academic Press.

Neelami, R., Choi, H., and Baraniuk, R. (1999). Wavelet-domain regularized deconvolution for ill-conditioned systems. Technical report, Deparment of Electrical and Computer Engineering, Rice University.

Strang, G. (1986). *Introduction to Applied Mathematics*. Wellesley-Cambridge Press.

Unser, M. and Blu, T. (1999). Construction of fractional spline wavelet bases. Technical report, Biomedical Imaging Group, Swiss Federal Institute of Technology, Lausanne.